

CONTEXT VECTOR GENERATION AND RETRIEVAL

ABSTRACT OF THE DISCLOSURE

A system and method for generating context vectors for use in storage and retrieval of documents and other information items. Context vectors represent conceptual relationships among information items by quantitative means. A neural network operates on a training corpus of records to develop relationship-based context vectors based on word proximity and co-importance using a technique of "windowed co-occurrence".

Relationships among context vectors are deterministic, so that a context vector set has one logical solution, although it may have a plurality of physical solutions. No human knowledge, thesaurus, synonym list, knowledge base, or conceptual hierarchy, is required. Summary vectors of records may be clustered to reduce searching time, by forming a tree of clustered nodes. Once the context vectors are determined, records may be retrieved using a query interface that allows a user to specify content terms, Boolean terms, and/or document feedback. The present invention further facilitates visualization of textual information by translating context vectors into visual and graphical representations. Thus, a user can explore visual representations of meaning, and can apply human visual pattern recognition skills to document searches.